Where to open a Spanish restaurant in Madrid

Project Report

SEPTEMBER 2019

PREPARED FOR
IBM Data Science Professional Certificate
Applied Data Science Capstone
Peer Graded Assignment: Capstone Project – The Battle of the Neighborhoods
https://github.com/mnlcsty/Coursera_Capstone

PREPARED BY
Manuel Costoya Ramos
manuel.costoya.ramos@gmail.com
https://www.linkedin.com/in/manuel-costoya-ramos

Table of Contents

Business Problem

Madrid is a big city. According to Wikipedia it is the third most populous city in the European Union behind Berlin and London. It has 131 boroughs grouped in 21 districts. It is a lively city where locals like to eat out, dine out, and in general go out. And it is now a popular travel destination for tourists from across the globe, so much so that in its September 2018 issue Time Out magazine named one of its neighborhoods “the coolest neighborhood in the world”.

Any time of the year, any day of the week, the Centro, Retiro and Salamanca districts are swarming with tourists and locals. The remaining districts inside the M-30 inner beltway (Arganzuela, Chamberí, Chamartín and Tetuán) are also very dynamic but the tourist presence is traditionally somewhat lower.

With so many locals and visitors eating out and dining out, it would seem that restaurants must be full all the time. And judging from the table occupancy of pavement cafés, that would seem to be the case. But the tables inside the restaurant frequently tell a very different story. More often than not they are all empty. Get a table in a pavement café if you are lucky enough, and you will realize that many people sit there for hours in front of a up of coffe, a beer or a bottle of water.

Complicating things further, a “restaurant bubble” has been growing in Madrid for some years, with more and more restaurants opening and closing every year. Popular gastronomy critic El Comidista dedicates this article to the problem.

As it happens the restaurant business is as hard in Madrid as anywhere else. A lot of people have lost money, gotten into debt, gone bankrupt and/or seriously damaged their family relationships after opening a restaurant attracted by the apparent business simplicity and movement of people, to find soon after how difficult the business really is.

The restaurant industry and its analysts have long studied the reasons why restaurants fail. Poor choice of location is among the most important, as discussed in the Literature Review section, and Data Science can help mitigate the associated risks.

This report is directed at anyone facing the choice of location for a new Spanish restaurant in Madrid.

Literature Review

Among the many articles and blog posts about the reasons for restaurant failure that can be found online, the one that stands out is “Why Restaurants Fail” by H.G. Parsa, John T. Self, David Njite and Tiffany King from Cornell University. In one of the sections the authors list the following elements of restaurant failure: lack of documented strategy, seat-of-the-pants management, management operations by “putting out fires”, focusing on only one aspect of the business, poor choice of location, restaurant concept and location mismatch, insufficient start-up or operational capital, lack of business experience or knowledge of restaurant operations, poor communication with customers, negative

customer perception of value, inability to mantain operational standards, loss of authenticity (for ethnic restaurants), becoming everything to everyone, underestimating the competition, lack of owner commitment due to family demands, lack of operational performance evaluation systems, frequent changes in management, tardy establishment of vision and mission statement, failure to maintain management flexibility and innovation, noncontrollable external factors, and entrepreneurial incompetence. Restaurant density is also discussed as an ambivalent factor: it can be beneficial to a point as it attracts more traffic, but it is strongly correlated with restaurant turnover.

The article conclusions seem as valid today as they were back in 2005 when it was written with one significant addition: social networking as a distinct and all-important form of customer relationship management.

Broadly speaking, the reasons for restaurant failure fall into one of the following categories:

- Lack of owner commitment, business knowledge and/or experience
- Insufficient capital
- Poor management, customer relationship marketing and/or social networking
- Loss of concept
- Poor choice of location

Of the above categories for restaurant failure , the one Data Science can help with is the choice of location. A model can be built to cluster all posible locations based on restaurant density, placement-related capital requirements, and direct competition—or suitable proxies. The best cluster(s) can then be chosen based on the effect of restaurant density and competition on failure risk and the owner's concept, decision criteria and access to capital. Depending on how the model features are distributed, one or several equally attractive locations may stand out as the best choices within the chosen clusters, or additional input by the owner may be required in order to make a choice.

Data

Madrid neighborhood list is available on the followingG Wikipedia page:
https://es.wikipedia.org/wiki/Anexo:Barrios_administrativos_de_Madrid.
The page will be scraped to extract for each neighborhood: district, neighborhood number, neighborhood, area and link to the neighborhood Wikipedia page.

The Wikipedia page for each neighborhood will in turn be scraped to extract the neighborhood latitude and longitude. For example, the page corresponding to Embajadores neighborhood is:
https://es.wikipedia.org/wiki/Embajadores_(Madrid).

The FourSquare ‘explore’ endpoint will be used to extract the most popular nearby venues for each neighborhood. The data will be used to compute neigborhood popularity and direct competitive pressure, both model features. The choice of the ‘explore’ over the ‘search’ endpoint will be justified in the Methodology section.

Finally, the 2nd hand housing prices for each neighborhood will be obtained from the following file published by the City Council:
https://www.madrid.es/UnidadesDescentralizadas/UDCEstadistica/Nuevaweb/Edificaci%C3%B3n%20y%20Vivienda/Mercado%20de%20la%20Vivienda/Precios%20de%20la%20Vivienda/Distritos/E3320219.xls.
The choice of this price index as a model feature will also be justified in the Methodology section.